

09898659-070301

Title: GENE CONTROLLING
FRUIT SIZE AND CELL
DIVISION IN PLANTS

Inventors: Steven D. Tanksley

Docket No.: 19603/3211 (CRF D-2594A)

GENE CONTROLLING FRUIT SIZE AND CELL DIVISION IN PLANTS

[0001] This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/215,824, filed July 5, 2000.

- 5 [0002] This invention was developed with government funding by the United States Department of Agriculture Grant No. 97-35300-4384; National Science Foundation Grant No. DBI-9872617; and Binational Agricultural Research and Development Fund No. US 2427-94. The U.S. Government may have certain rights.

10

FIELD OF THE INVENTION

[0003] The present invention relates to the identification of a gene which controls fruit size and/or cell division in plants, the proteins encoded by that gene, and uses thereof.

15

BACKGROUND OF THE INVENTION

- [0004] In natural populations, most phenotypic variation is continuous and effected by alleles at multiple loci. Although this quantitative variation fuels evolutionary change and has been exploited in the domestication and genetic improvement of plants and animals, the identification and isolation of the genes underlying this variation has been difficult.
- 20

- [0005] The most conspicuous and, perhaps, most important quantitative traits in plant agriculture are those associated with domestication (Doebley et al., "Genetic and Morphological Analysis of a Maize-Teosinte F₂ Population: Implications for the Origin of Maize," PNAS 87: 9888-9892 (1990)). Key adaptations to survival in the wild were dramatically modified by early humans; fruit-bearing crop plants are a prime example. Dramatic and relatively rapid changes in fruit size have accompanied the domestication of virtually all fruit-bearing crop species, including tomato, watermelon, apple, banana, grape, berries and a vast assortment of other tropical, subtropical, and temperate species (J.
- 25
- 30

Smartt et al., Evolution of Crop Plants (Longman Group, United Kingdom, (1995)). These changes have benefited mankind but have often been at the expense of the plant's seed production, dispersal, and survival under natural conditions. The progenitor of domesticated tomato (*Lycopersicon esculentum* Mill.) most likely had fruit less than 1 cm in diameter and only a few grams in weight (Rick, C. M., "Tomato," Scientific American 239:76 (1978)). Such fruit were large enough to contain hundreds of seeds and yet small enough to be dispersed by small rodents or birds. In contrast, modern tomatoes can weigh as much as 1,000 grams and can exceed 150 cm in diameter. While it is known that the transition from small to large fruit occurred numerous times during the domestication of crop plants (J. Smartt, et al. Evolution of Crop Plants (Longman Group, United Kingdom, (1995)) and that it is quantitatively controlled (Paterson et al., "Mendelian Factors Underlying Quantitative Traits in Tomato: Comparison Across Species, Generations, and Environments," Genetics 127(1):181-97 (1991)), the molecular basis of this transition has thus far been unknown.

[0006] Using the approach of quantitative trait locus (QTL) mapping (Lander et al., "Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps," Genetics 121(1):185-99 (1989) published erratum appears in Genetics 136 (2):705 (1994); Tanksley S.D., "Mapping Polygenes," Annu Rev Genet 27:205-33 (1993)), most of the loci involved in the evolution and domestication of tomato from small berries to large fruit have been genetically mapped (Grandillo et al., "Identifying the Loci Responsible for Natural Variation in Fruit Size and Shape in Tomato," Theor. Appl. Gen. 99:978 (1999)). One of these QTLs, *fw2.2*, appears to have been responsible for a key transition during domestication: all wild *Lycopersicon* species examined thus far contain small fruit alleles at this locus whereas modern cultivars have large fruit alleles (Alpert et al., "FW-2.2 – A Major QTL Controlling Fruit Weight Is Common to Both Red-Fruited and Green-Fruited Tomato Species," Theor. Appl. Gen. 91: 994 (1995)). What is needed to further the current understanding of the genetic regulation of fruit size in plants is the identification of the nucleic acid sequence of the *fw2.2* gene and of the protein product encoded by the cDNA of that gene.

[0007] The present invention is directed to achieving these objectives.

SUMMARY OF THE INVENTION

[0008] The present invention relates to an isolated nucleic acid molecule encoding a protein which regulates fruit size and/or cell division in plants.

[0009] The present invention also relates to an isolated protein which
5 regulate fruit size and/or cell division in plants.

[0010] The present invention also relates to a method of regulating fruit size in plants by transforming a plant with a nucleic acid molecule of the present invention under conditions effective to regulate fruit size in the plant.

[0011] The present invention also relates to a method of regulating cell
10 division in plants by transforming a plant with a nucleic acid molecule of the present invention under conditions effective to regulate cell division in the plant.

[0012] The present invention provides an important advance in the study of morphogenesis in plants, and provides new opportunities for understanding and utilizing natural variation. In particular, a greater understanding of the genetic
15 regulation of fruit size and/or cell division in plants provides a means for the generation of agronomically superior crops through genetic manipulation.

BRIEF DESCRIPTION OF THE DRAWINGS

[0013] Figure 1A shows the fruit size extremes in the genus *Lycopersicon*.
20 On the left is a fruit from the wild tomato species *L. pimpinellifolium*, which, like all other wild tomato species, bears very small fruit. On the right is a fruit from *L. esculentum* cv Giant Red, bred to produce extremely large tomatoes. Figure 1B shows the phenotypic effect of the *fw2.2* transgene in the cultivar Mogeor. Fruit are from R1 progeny of #fw107 segregating for the presence (+) and absence (-)
25 of *cos50* containing the small fruit allele.

[0014] Figures 2A-C show the high-resolution mapping of the *fw2.2* QTL. Figure 2A shows the location of *fw2.2* on tomato chromosome 2 in a cross between *L. esculentum* and a nearly isogenic line (NIL) containing a small introgression (grey area) from *L. pennellii*. Figure 2B shows a contig of the *fw2.2*

candidate region, delimited by recombination events at XO31 and XO33. Figure 2C shows a sequence analysis of the cos50 transgene.

[0015] Figures 3A-E show the reverse transcriptase and histological analyses of the large and small-fruited NILs, TA1143 and TA1144, respectively.

5 [0016] Figures 4A-B show a CLUSTALW alignment of LpORFX (*L.pennellii*, AF261775) and LcORFX (*L.esculentum*, AF261774) with seven representatives of 26 matched from the Genbank Expressed Sequence Tag ("EST") and nucleotide databases and the contigs assembled from the TIGR tomato EST database. Sequences begin on Figure 4A and continue onto Figure
10 4B.

[0017] Figure 5A shows the secondary structure analysis of the predicted ORFX protein, which indicates that ORFX is a soluble protein with α/β type secondary structure. Figure 5B shows the threading program LOOPP analysis which assigns ORFX to the fold of 6q21, domain A, and gives the Z-scores for
15 global and local alignments.

DETAILED DESCRIPTION OF THE INVENTION

[0018] The present invention relates to an isolated nucleic acid molecule which regulates fruit size and/or cell division in plants.

20 [0019] One embodiment of the nucleic acid molecule of the present invention is a nucleic acid molecule that encodes a protein which reduces fruit size and/or cell division in plants. An example of such a nucleic acid molecule is isolated from the small-fruited tomato *Lycopersicon pennellii* which has a nucleotide sequence corresponding to SEQ. ID. No. 1 as follows:

atgtatccaa cggtaggata taatctaggt ctaatgaaac aaccttatgt tcttcotcac 60
 tatgtatctg ccccgggcac caccacggcg cggtgggtcaa ctggtctttt tcactgtttt 120
 gatgaccttg ctaactgttt agttactagt gtttgcctt gtatcacct tggacagatt 180
 tctgaaatac taaacaaagg aacaacttca tgtgggagta gaggtgcatt atattgtttg 240
 ctgggaactga caggattgcc tagcctatat tctgtcttct acaggtctaa aatgaggggg 300
 caatatgac tggaagaggc acctgtgtt gattgtcttg tacatgtatt ctgtgaacct 360
 tgtgtctctt gccaaagaata cagagagctt aagaaccgtg gctttgatat gggaaatagg 420
 tggcaagcta atatggatag acaaagccgg ggagttacca tgccccctta tcatgcaggc 480
 atgaccaggt ga 492

The nucleotide sequence of SEQ. ID. No. 1 encodes a protein, LpORFX, having an amino acid sequence corresponding to SEQ. ID. No. 2, as follows:

Met Tyr Pro Thr Val Gly Tyr Asn Leu Gly Leu Met Lys Gln Pro Tyr
 1 5 10 15

Val Pro Pro His Tyr Val Ser Ala Pro Gly Thr Thr Thr Ala Arg Trp
 20 25 30

Ser Thr Gly Leu Cys His Cys Phe Asp Asp Pro Ala Asn Cys Leu Val
 35 40 45

Thr Ser Val Cys Pro Cys Ile Thr Phe Gly Gln Ile Ser Glu Ile Leu
 50 55 60

Asn Lys Gly Thr Thr Ser Cys Gly Ser Arg Gly Ala Leu Tyr Cys Leu
 65 70 75 80

Leu Gly Leu Thr Gly Leu Pro Ser Leu Tyr Ser Cys Phe Tyr Arg Ser
 85 90 95

Lys Met Arg Gly Gln Tyr Asp Leu Glu Glu Ala Pro Cys Val Asp Cys
 100 105 110

Leu Val His Val Phe Cys Glu Pro Cys Ala Leu Cys Gln Glu Tyr Arg
 115 120 125

Glu Leu Lys Asn Arg Gly Phe Asp Met Gly Ile Gly Trp Gln Ala Asn
 130 135 140

Met Thr Arg
163

atgtatcaaa	cggtaggata	taatccaggt	ccaatgaac	aaccttatgt	tctctctcac	60
tatgtatctg	cccccggaac	caccacggcg	cgggtgtcga	ctgtgtcttg	tcatgtgttt	120
gatgaccttg	ctaaactgttt	agttactagt	gtttgcacct	gtatcacctt	tgagacagat	180
ctcgaataac	taaaccaagg	aaccaactta	gttgggagta	gaggtgcatt	atattgtttg	240
ctggagtga	caggatgcgc	tacgctatat	tctctgtctt	acaggctctaa	aatgaggggg	300
caatatgac	tggaaagagc	acottgtgtt	gattgtcttg	tacatgtatt	ctgtgaacct	360
tgtgctcttt	gccaaagaat	cagagagctt	aagaaccgtg	gctttgatat	gggaataggg	420
tggcaagct	atatgtagat	acaaagccga	ggagttacca	tgccccctta	tcatgcaggc	480
atgccaggt	ga					492

Met Tyr Gln Thr Val Gly Tyr Asn Pro Gly Pro Met Lys Gln Pro Tyr
1 5 10 15

Ser Thr Gly Leu Cys His Cys Phe Asp Asp Pro Ala Asn Cys Leu Val
35 40 45

Asn Lys Gly Thr Thr Ser Cys Gly Ser Arg Gly Ala Leu Tyr Cys Leu
35 65 70 75 80

Leu Gly Leu Thr Gly Leu Pro Ser Leu Tyr Ser Cys Phe Tyr Arg Ser
85 90 95

Lys Met Arg Gly Gln Tyr Asp Leu Glu Glu Ala Pro Cys Val Asp Cys
 100 105 110

5 Leu Val His Val Phe Cys Glu Pro Cys Ala Leu Cys Gln Glu Tyr Arg
 115 120 125

Glu Leu Lys Asn Arg Gly Phe Asp Met Gly Ile Gly Trp Gln Ala Asn
 130 135 140

10

Met Asp Arg Gln Ser Arg Gly Val Thr Met Pro Pro Tyr His Ala Gly
 145 150 155 160

Met Thr Arg

- 15 **[0021]** Sequence analysis of the nucleic acid molecule of the present invention, known herein as *ORFX*, and described in greater detail below, revealed that it contains two introns and encodes a 163 amino acid polypeptide of approximately 22kDa. Protein secondary structure prediction algorithms (Rost et al., "Combining Evolutionary Information and Neural Networks To Predict Protein Secondary Structure," *Proteins* 19(1):55-72 (1994), which is hereby incorporated by reference in its entirety) suggest the *ORFX* protein has two to three hydrophobic β -strands, separated by hydrophilic turn domains, with a possible single helix near the carboxy-terminus, suggesting an overall β -sheet or mixed α - β structure. The presence of twelve highly conserved cysteine residues indicates possible zinc-finger-like domains (and thus potential interaction of the protein with DNA), but their distribution does not fit the pattern of previously characterized zinc-fingers (Struhl K., "Helix-Turn-Helix, Zinc-Finger, and Leucine-Zipper Motifs for Eukaryotic Transcriptional Regulatory Proteins," *Trends Biochem Sci* 14(4):137-40 (1989), which is hereby incorporated by reference in its entirety). The first forty amino-terminal residues are relatively hydrophilic and unstructured and are poorly conserved between putative homologs. Additional sequence analysis reveals no significant similarity to known protein motifs (BLOCKS+) (Henikoff et al., "Protein Family Classification Based On Searching A Database of Blocks," *Genomics* 1:19(1):97-107 (1994), which is hereby incorporated by reference in its entirety) or protein localization
- 20
- 25
- 30
- 35

000055950000

signals (PSORT)(Nakai et al., "A Knowledge Base For Predicting Protein Localization Sites in Eukaryotic Cells," Genomics 14(4):897-911 (1992), which is hereby incorporated by reference in its entirety).

5 [0022] Also suitable as a nucleic acid molecule according to the present invention is an isolated nucleic acid molecule encoding a protein which controls fruit size and/or plant cell division, wherein the nucleic acid selectively hybridizes to the nucleotide sequence of SEQ. ID. No. 1 or SEQ. ID. No. 3 under stringent conditions characterized by a hybridization buffer comprising 0.9M sodium citrate buffer at a temperature of 45°C.

10 [0023] Fragments of the above proteins are also encompassed by the present invention. Suitable fragments can be produced by several means. In the first, subclones of the gene encoding the protein of the present invention are produced by conventional molecular genetic manipulation by subcloning gene fragments. The subclones then are expressed *in vitro* or *in vivo* in bacterial cells
15 to yield a smaller protein or peptide.

[0024] In another approach, based on knowledge of the primary structure of the protein of the present invention, fragments of the gene of the present invention may be synthesized by using the PCR technique together with specific sets of primers chosen to represent particular portions of the protein. These then
20 would be cloned into an appropriate vector for increased expression of an accessory peptide or protein.

[0025] Chemical synthesis can also be used to make suitable fragments. Such a synthesis is carried out using known amino acid sequences for the protein of the present invention. These fragments can then be separated by conventional
25 procedures (e.g., chromatography, SDS-PAGE) and used in the methods of the present invention.

[0026] Variants may also (or alternatively) be prepared by, for example, the deletion or addition of amino acids that have minimal influence on the properties, secondary structure, and hydrophobic nature of the polypeptide. For
30 example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally

directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification, or identification of the polypeptide.

[0027] The present invention also relates to an expression vector containing a DNA molecule encoded by the nucleic acid molecules of the present invention. The nucleic acid molecules of the present invention may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. In preparing a DNA vector for expression, the various DNA sequences may normally be inserted or substituted into a bacterial plasmid. Any convenient plasmid may be employed, which will be characterized by having a bacterial replication system, a marker which allows for selection in a bacterium, and generally one or more unique, conveniently located restriction sites. Numerous plasmids, referred to as transformation vectors, are available for plant transformation. The selection of a vector will depend on the preferred transformation technique and target species for transformation. A variety of vectors are available for stable transformation using *Agrobacterium tumefaciens*, a soilborne bacterium that causes crown gall. Crown gall are characterized by tumors or galls that develop on the lower stem and main roots of the infected plant. These tumors are due to the transfer and incorporation of part of the bacterium plasmid DNA into the plant chromosomal DNA. This transfer DNA (T-DNA) is expressed along with the normal genes of the plant cell. The plasmid DNA, pTI, or Ti-DNA, for "tumor inducing plasmid," contains the *vir* genes necessary for movement of the T-DNA into the plant. The T-DNA carries genes that encode proteins involved in the biosynthesis of plant regulatory factors, and bacterial nutrients (opines). The T-DNA is delimited by two 25 bp imperfect direct repeat sequences called the "border sequences." By removing the oncogene and opine genes, and replacing them with a gene of interest, it is possible to transfer foreign DNA into the plant without the formation of tumors or the multiplication of *Agrobacterium tumefaciens* (Fraley, et al., "Expression of Bacterial Genes in Plant Cells," Proc. Nat'l Acad. Sci., 80:4803-07 (1983), which is hereby incorporated by reference in its entirety).

[0028] Further improvement of this technique led to the development of the binary vector system (Bevan, M., "Binary *Agrobacterium* Vectors for Plant Transformation," Nucleic Acids Res. 12:8711-21 (1984), which is hereby incorporated by reference in its entirety). In this system, all the T-DNA sequences (including the borders) are removed from the pTi, and a second vector containing T-DNA is introduced into *Agrobacterium tumefaciens*. This second vector has the advantage of being replicable in *E. coli* as well as *A. tumefaciens*, and contains a multiclonal site that facilitates the cloning of a transgene. An example of a commonly used vector is pBin19 (Frisch, et al., "Complete Sequence of the Binary Vector Bin19," Plant Molec. Biol. 27:405-409 (1995), which is hereby incorporated by reference in its entirety). Any appropriate vectors now known or later described for genetic transformation are suitable for use with the present invention.

[0029] U.S. Patent No. 4,237,224 issued to Cohen and Boyer, which is hereby incorporated by reference in its entirety, describes the production of expression systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase. These recombinant plasmids are then introduced by means of transformation and replicated in unicellular cultures including prokaryotic organisms and eukaryotic cells grown in tissue culture.

[0030] In one aspect of the present invention, the nucleic acid molecules of the present invention are individually incorporated into an appropriate vector in the sense direction, such that the open reading frame is properly oriented for the expression of the encoded protein under control of a promoter of choice.

[0031] Certain "control elements" or "regulatory sequences" are also incorporated into the vector-construct. These include non-translated regions of the vector, promoters, and other 5' or 3' untranslated regions which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used.

[0032] A constitutive promoter is a promoter that directs expression of a gene throughout the development and life of an organism. Examples of some

5
10

15
20
25
30

Death, Plant J. 14(2):247-57 (1998), which are hereby incorporated by reference in their entirety). In addition, inducible promoters include promoters that function in a tissue specific manner to regulate the gene of interest within selected tissues of the plant. Examples of such tissue specific promoters include seed, flower, or root specific promoters as are well known in the field (U.S. Patent No. 5,750,385 to Shewmaker et al., which is hereby incorporated by reference in its entirety).

[0034] The DNA construct of the present invention also includes an operable 3' regulatory region, selected from among those which are capable of providing correct transcription termination and polyadenylation of mRNA for expression in the host cell of choice, operably linked to a DNA molecule which encodes for a protein of choice. A number of 3' regulatory regions are known to be operable in plants. Exemplary 3' regulatory regions include, without limitation, the nopaline synthase 3' regulatory region (Fraley, et al., "Expression of Bacterial Genes in Plant Cells," Proc. Nat'l Acad. Sci. USA 80:4803-07 (1983), which is hereby incorporated by reference in its entirety) and the cauliflower mosaic virus 3' regulatory region (Odell, et al., "Identification of DNA Sequences Required for Activity of the Cauliflower Mosaic Virus 35S Promoter," Nature 313(6005):810-812 (1985), which is hereby incorporated by reference in its entirety). Virtually any 3' regulatory region known to be operable in plants would suffice for proper expression of the coding sequence of the DNA construct of the present invention.

[0035] The vector of choice, promoter, and an appropriate 3' regulatory region can be ligated together to produce the plasmid, or DNA construct, of the present invention using well known molecular cloning techniques as described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Press, NY (1989), and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y., which are hereby incorporated by reference in their entirety.

[0036] A further aspect of the present invention is a host cell which includes a DNA construct of the present invention. As described more fully hereinafter, the recombinant host cell can be either a bacterial cell (e.g., *Agrobacterium*), a virus, or a plant cell. In the case of recombinant plant cells, it

is preferable that the DNA construct is stably inserted into the genome of the recombinant plant cell.

[0037] The DNA construct can be incorporated into cells using conventional recombinant DNA technology. Generally, this involves inserting the DNA construct into an expression vector or system to which it is heterologous (i.e., not normally present). As described above, the DNA construct contains the necessary elements for the transcription and translation of the heterologous DNA molecule in plant cells.

[0038] Once the DNA construct of the present invention has been prepared, it is ready to be incorporated into a host cell. Recombinant molecules can be introduced into cells via transformation, particularly transduction, conjugation, mobilization, or electroporation. Suitable host cells include, but are not limited to, bacteria, virus, yeast, mammalian cells, insect, plant, and the like. Preferably the host cells are either a bacterial cell or a plant cell.

[0039] Accordingly, another aspect of the present invention relates to a method of making a recombinant plant cell. Basically, this method is carried out by transforming a plant cell with a DNA construct of the present invention under conditions effective to yield transcription of the DNA molecule in response to the promoter. Methods of transformation may result in transient or stable expression of the DNA under control of the promoter. Preferably, the DNA construct of the present invention is stably inserted into the genome of the recombinant plant cell as a result of the transformation, although transient expression can serve an important purpose, particularly when the plant under investigation is slow-growing.

[0040] One approach to transforming plant cells with a DNA construct of the present invention is particle bombardment (also known as biolistic transformation) of the host cell. This can be accomplished in one of several ways. The first involves propelling inert or biologically active particles at cells. This technique is disclosed in U.S. Patent Nos. 4,945,050, 5,036,006, and 5,100,792, all to Sanford, et al., which are hereby incorporated by reference in their entirety. Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and

embodiment of the present invention transformants are generated using the method of Frary et al, "An Examination of Factors Affecting the Efficiency of Agrobacterium-Mediated Transformation of Tomato," Plant Cell Reports 16: 235 (1996), which is hereby incorporated by reference in its entirety, to transform seedling explants.

[0043] Plant tissues suitable for transformation include, but are not limited to, floral buds, leaf tissue, root tissue, meristems, zygotic and somatic embryos, megasporos, and anthers.

[0044] After transformation, the transformed plant cells can be selected and regenerated. Preferably, transformed cells are first identified using a selection marker simultaneously introduced into the host cells along with the DNA construct of the present invention. The most widely used reporter gene for gene fusion experiments has been *uidA*, a gene from *Escherichia coli* that encodes the β -glucuronidase protein, also known as GUS (Jefferson et al., "GUS Fusions: β Glucuronidase as a Sensitive and Versatile Gene Fusion Marker in Higher Plants," EMBO Journal 6:3901-3907 (1987), which is hereby incorporated by reference in its entirety). GUS is a 68.2 kd protein that acts as a tetramer in its native form. It does not require cofactors or special ionic conditions, although it can be inhibited by divalent cations like Cu^{2+} or Zn^{2+} . GUS is active in the presence of thiol reducing agents like β -mercaptoethanol or dithiothreitol (DTT).

[0045] In order to evaluate GUS activity, several substrates are available. The most commonly used are 5 bromo-4 chloro-3 indolyl glucuronide (X-Gluc) and 4 methyl-umbelliferyl-glucuronide (MUG). The reaction with X-Gluc generates a blue color that is useful in histochemical detection of the gene activity. For quantification purposes, MUG is preferred, because the umbelliferyl radical emits fluorescence under UV stimulation, thus providing better sensitivity and easy measurement by fluorometry (Jefferson et al., "GUS Fusions: β Glucuronidase as a Sensitive and Versatile Gene Fusion Marker in Higher Plants," EMBO Journal 6:3901-3907 (1987), which is hereby incorporated by reference in its entirety).

15

25

30

Plants, Acad. Press, Orlando, Vol. I, 1984, and Vol. III (1986), which are hereby incorporated by reference in their entirety.

[0049] It is known that practically all plants can be regenerated from cultured cells or tissues. This includes, but is not limited to, all major crop plants, such as rice, wheat, barley, rye, cotton, sunflower, peanut, corn, potato, sweet potato, bean, pea, chicory, lettuce, endive, cabbage, cauliflower, broccoli, turnip, radish, spinach, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon, strawberry, grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum, and sugarcane. Transgenic ornamental plants, such as *Arabidopsis thaliana*, *Saintpaulia*, petunia, pelargonium, poinsettia, chrysanthemum, carnation, and zinnia, can also be produced which harbor the nucleic acid molecules of the present invention.

[0050] After a DNA construct of the present invention is stably incorporated in transgenic plants, it can be transferred to other plants by sexual crossing or by preparing cultivars. With respect to sexual crossing, any of a number of standard breeding techniques can be used depending upon the species to be crossed. Cultivars can be propagated in accord with common agricultural procedures known to those in the field. Alternatively, transgenic seeds are recovered from the transgenic plants. The seeds can then be planted in the soil and cultivated using conventional procedures to produce transgenic plants.

[0051] Another aspect of the present invention relates to a method of regulating fruit size in a plant. This involves transforming a host which is a plant cell with the expression vector containing a nucleic acid of the present invention, under conditions effective to regulate fruit size in the plant. This method is carried out by transforming a plant cell with a construct of the present invention. In one embodiment of this aspect, the construct of the present invention is cloned into the expression vector in proper sense orientation and correct reading frame. Transgenic plants are produced as described above, which exhibit a fruit size that is modified from its normal phenotype. The phenotypic effect is to reduce fruit size when the construct contains a nucleic acid molecule having SEQ. ID. No. 1. When a nucleic acid molecule having SEQ. ID. No. 3 is used in the construct the phenotypic effect will be to increase fruit size of the plant. Preferably, the

construct of the present invention is stably inserted into the genome of the recombinant plant cell as a result of the transformation.

- [0052] Another aspect of the present invention relates to a method of regulating cell division in plants. This involves transforming a plant, as described
- 5 above, with the nucleic acid molecules of the present invention, under conditions effective to regulate cell division in a plant. This involves transforming a plant cell with a construct of the present invention, as describe above. This method may be carried out on a variety of plant tissues, as the regulation of cell division has numerous applications. For example, cell division in carpels (which develop in
- 10 fruit), sepals, and styles may be increased or decreased relative to the native phenotype of the plant depending on whether the nucleic acid molecule corresponding to SEQ. ID. No. 1 or SEQ. ID. No. 3 of the present invention is the transgene. If transformation is carried out with the nucleic acid molecule corresponding to SEQ. ID. No. 1 of the present invention, decreased cell division
- 15 will occur in the transgenic plant, with plant organs, including, but not limited to, carpels, styles, and sepals of the transgenic plant. Conversely, cell division will be increased in plants transformed with SEQ. ID. No. 3 of the present invention, producing larger organs in the plant. This method of regulating cell division can be applied to many types of plants. This includes, but is not limited to, all major
- 20 crop plants, such as rice, wheat, barley, rye, cotton, sunflower, peanut, corn, potato, sweet potato, bean, pea, chicory, lettuce, endive, cabbage, cauliflower, broccoli, turnip, radish, spinach, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon, strawberry, grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum, and sugarcane.
- 25 Ornamental plants, such as *Arabidopsis thaliana*, *Saintpaulia*, petunia, pelargonium, poinsettia, chrysanthemum, carnation, and zinnia, can also be used with this method of regulating cell division.

EXAMPLES

Example 1 - Genetic Complementation with *fw2.2*

- [0053]** A yeast artificial chromosome (YAC) containing the QTL *fw2.2* was isolated and used to screen a cDNA library constructed from the small-fruited genotype, *L. pennellii* LA716. Approximately 100 positive cDNA clones were identified that represent four unique transcripts: cDNA27, cDNA38, cDNA44 and cDNA70, that were derived from genes in the *fw2.2* YAC contig. The four cDNAs were then used to screen a cosmid library of *L. pennellii* genomic DNA that was constructed in the binary cosmid transformation vector TDNA 04541. For the cosmid library screen, the cDNAs were sequenced and specific primers were designed for a PCR-based screen of the pooled library. Positive pools were then plated, lifted, and probed with the corresponding cDNA. Four positive, non-overlapping cosmids (cos50, cos62, cos69, and cos84) were identified, one corresponding to each unique transcript. These four cosmid clones were assembled into a physical contig of the *fw2.2* region using the Long Template PCR System, using manufacturer's directions (Boehringer Mannheim, Indianapolis, IN). Cosmids cos50, cos62, cos69, and cos84 were used for genetic complementation analysis in transgenic plants.
- [0054]** The constructs were transformed into two tomato cultivars, Mogeor (fresh market-type) and TA496 (processing-type) using the method of Frary et al., "An Examination of Factors Affecting the Efficiency of Agrobacterium-Mediated Transformation of Tomato," Plant Cell Reports 16: 235 (1996), which is hereby incorporated by reference in its entirety. Both tomato lines carry the partially recessive large fruit allele of *fw2.2*. As *fw2.2* is a quantitative trait locus and the *L. pennellii* allele is only partially dominant, the primary transformants (R0), which are hemizygous for the transgene, were self-pollinated to obtain segregating R1 progeny. Putative transformants were assayed using PCR and Southern hybridization for the neomycin phosphatase II (*nptII*) selectable marker gene that every construct carried.
- [0055]** Figure 1A shows the fruit size extremes in the genus *Lycopersicon*. In plants containing the transgene of the present invention, a statistically

significant reduction in fruit weight indicated that the plants were carrying the small fruit allele of *fw2.2* and that complementation had been achieved. This result was only observed in the R1 progeny of primary transformants #fw71 and #fw107 both of which carried *cos50*. Figure 1B shows the phenotypic effect of the *fw2.2* transgene in the cultivar Mogeor. Fruit are from R1 progeny of the #fw107 segregating for the presence (+) of *cos50*, shown on the right panel of Figure 1B, and the absence (-) of *cos50*, shown in the left panel of Figure 1B. Table 1 gives the average fruit weight and seed numbers for R1 progeny of several primary transformants. Unless otherwise noted, progeny are from independent R0 plants. Numbers in parentheses are the numbers of R1 individuals tested.

0000559.070201

Table 1.

Cosmid	Cultivar	R0 plant #	Average fruit weight (g)		P-value	Average seed number		P-value
			+ transgene	-transgene		+ transgene	-transgene	
50*	TA496	fw71	41.6 (18)	56.4 (7)	<0.0001	32.6 (18)	28.3 (7)	0.40
50*	TA496	fw71	47.7 (23)	68.1 (12)	<0.0001	31.4 (23)	27.4 (12)	0.44
50	Mogeor	fw107	25.4 (21)	40.9 (7)	<0.0001	24.1 (21)	28.2 (7)	0.34
62	Mogeor	fw59	46.5 (18)	48.0 (9)	0.70	36.1 (18)	36.5 (9)	0.94
62	TA496	fw70	51.0 (21)	51.3 (3)	0.94	28.3 (21)	39.8 (3)	0.04
69	Mogeor	fw51	50.0 (14)	51.7 (10)	0.58	29.8 (14)	34.8 (10)	0.15
84	Mogeor	fw95	49.4 (18)	47.9 (5)	0.71	33.0 (18)	35.5 (5)	0.62

* R1 progeny of the same primary transformant.

[0056] Seed number is included in the analysis, because reduced fertility, as evidenced by reduced seed per fruit, can decrease fruit size. Thus, these data show that the change in fruit size associated with *cos50* is not a byproduct of reduced fertility.

- 5 [0057] The fact that the two complementing transformation events are independent and in different tomato lines (TA496 and Mogeor) indicates that the *cos50* transgene functions similarly in different genetic backgrounds and genomic locations. Thus, the progeny of plants #fw71 and #fw107 show that *fw2.2* is contained within *cos50*.
- 10 [0058] Most QTL alleles are not fully dominant or recessive (Lander et al., "Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps," Genetics 121(1):185-99 (1989), which is hereby incorporated by reference in its entirety). The small fruit *L. pennellii* allele for *fw2.2* is semi-dominant to the large fruit *L. esculentum* allele (Grandillo et al., "Identifying the
- 15 Loci Responsible for Natural Variation in Fruit Size and Shape in Tomato," Theor. Appl. Gen. 99:978 (1999), which is hereby incorporated by reference in its entirety). R2 progeny of #fw71 were used to calculate the gene action (d/a = dominance deviation/additivity; calculated as described in Grandillo et al., "Identifying the Loci Responsible for Natural Variation in Fruit Size and Shape in
- 20 Tomato," Theor. Appl. Gen. 99:978 (1999), which is hereby incorporated by reference in its entirety) of *cos50* in the transgenic plants. The transgene had a d/a of 0.51; in previous work using NILs, *fw2.2* had a d/a of 0.44. This similarity of gene action is consistent with the conclusion that the *cos50* transgene carries *fw2.2*.
- 25 **Example 2 - *fw2.2* Corresponds to *ORFX* and is Expressed in Pre-Anthesis Floral Organs**
- [0059] Figure 2A shows the location of *fw2.2* on tomato chromosome 2 in a cross between *L. esculentum* and a NIL containing a small introgression (gray area) from *L. pennellii*. Sequence analysis of *cos50* revealed two open reading
- 30 frames ("ORF"s), shown in Figure 2A: one corresponding to cDNA44, which was used to isolate *cos50*, and another 663 nucleotide (nt) gene, *ORFX*, for which no corresponding transcript was detected in the initial cDNA library screen. The

insert also contains a highly repetitive, AT-rich (80%) region of 1.4 kb. Previous mapping of *fw2.2* had identified a single recombination event which delimited the "right-most" end of the *fw2.2* candidate region (XO33, as described in Alpert et al., "FW-2.2 - A Major QTL Controlling Fruit Weight Is Common to Both Red-Fruited and Green-Fruited Tomato Species," Theor. Appl. Gen. 91: 994 (1995), which is hereby incorporated by reference in its entirety). Comparison of genomic DNA sequence from this recombinant plant with that of the two parental lines indicated that XO33 is within 43 to 80 nucleotides 5' from the end of *ORFX*, shown in Figure 2A. Because genetic mutation(s) causing change in fruit size must be to the left of XO33, cDNA44 cannot be involved and *ORFX* or an upstream region is the likely cause of the *fw2.2* QTL phenotype. Figure 2B shows the contig of the *fw2.2* candidate region, delimited by recombination events at XO31 and XO33. Arrows represent the four original candidate cDNAs (70, 27, 38, and 44, discussed in Example 1) and heavy horizontal bars are the four cosmids (cos62, 84, 69, and 50) isolated using these cDNAs as probes. The vertical lines are positions of RFLP or CAPs markers. Figure 2C is the sequence analysis of cos50, including the positions of cDNA44, *ORFX*, the region showing similarity to a *S-tuberosum* intergenic spacer (IGS), and the "right-most" recombination event, XO33.

[0060] *ORFX* is transcribed at levels too low to be detected through standard northern hybridization protocols in all pre-anthesis floral organs (petal, carpels, sepals, stamen) of both large and small fruited NILs; however, semi-quantitative reverse transcriptase analysis indicated that the highest levels were expressed in carpels. In addition, comparison of the relative levels of *ORFX* transcript in the carpels of the NILs showed significantly higher levels in the small-fruited NIL (TA1144) than in the large-fruited NIL (TA1143), as shown in Figure 3A. Figure 3A is a gel showing RT-PCR products for *ORFX* in various stages/organs. Stage I = 3 to 5 mm floral buds; Stage II = 5 mm to anthesis; Stage III = anthesis; 1 = sepals; 2 = petals; 3 = stamen; 4 = carpels; L = leaves. The observation of *ORFX* transcription in pre-anthesis carpels suggests that *fw2.2* exerts its effect early in development. To test this hypothesis, a comparison was made of the floral organs from the small and large fruited NILs. The results of

00000000-070001

this comparison are shown in Figures 3B-E. Top sections, Figure 3B and Figure 3C, display cortical cells from carpel septum. Bottom sections, Figure 3D and Figure 3E, display pericarp cells from carpel walls. Sections on the left, Figure 3B and Figure 3D, were derived from carpels of NIL homozygous for large fruit allele. Sections on right, Figure 3C and Figure 3E, were derived from carpels of NIL homozygous for small fruit allele. Carpels (which ultimately develop into fruit), styles, and sepals of the large-fruited NIL were already significantly heavier at anthesis ($p = 0.0007$, 0.001 , and 0.001 , respectively) than their counterparts in the small-fruited NIL. Stamen and petals showed no significant difference ($p = 0.63$ and 0.74 , respectively). Cell sizes at anthesis are similar ($p = 0.98$ and $p = 0.85$) in the NILs. Hence, carpels of large fruited genotypes contain more cells. Therefore, it was concluded that allelic variation at *ORFX* modulates fruit size at least in part by controlling carpel cell number prior to anthesis. TA1143 and TA1144 were not significantly different for cell size in either carpel walls (cells per $\text{mm}^2 = 17,600 \pm 700$ vs. $17,700 \pm 1000$; $p = 0.98$) or carpel septa (cells per $\text{mm}^2 = 10,100 \pm 500$ vs. $10,300 \pm 900$; $p = 0.85$) (statistical analysis based on 144 cell area counts from 48 sections). Carpels were fixed in 2.5% glutaraldehyde, 2% paraformaldehyde, 0.1 M Na cacodylate buffer, pH 6.8, and embedded in Spurr plastic. Bar represents $20\mu\text{M}$.

Example 3 - Sequence Analysis of ORFX

20 [0061] Total RNA was extracted with TRIzol reagent as described by the manufacturer (Gibco BRL, Grand Island, NY). First-strand cDNA was synthesized using SuperscriptTM RNaseH⁻ Reverse Transcriptase (Gibco BRL, Grand Island, NY) with the following primers:

25 B₂₆ primer, corresponding to SEQ. ID. No. 5, as follows:
5' GACTCGAGTCGACATCGA(dT)₁₇ 3';

B₂₅ primer, corresponding to SEQ.ID. No. 6, which was used for 3' RACE PCR to amplify *ORFX* transcript, as follows:

30 5' GACTCGAGTCGACATCGA 3';
and ORFX₂, corresponding to SEQ. ID. No. 7 as follows:
5' AAACAACCTTATGTTCTCCTCTCA 3'.

[0062] Nested PCR was carried out using primer B₂₅ (SEQ. ID. No. 6) and FW01, corresponding to SEQ. ID. No. 8, as follows:

5' GCCCTTGATCACCTTTGGA 3'.

[0063] The 5' RACE system (Gibco BRL, Grand Island, NY) was

- 5 employed to characterize the start of transcription of *ORFX*. Total RNA(5µg) was mixed with GSP₁ primer corresponding to SEQ. ID. No. 9, as follows:

5' GATGATTTCATTGATCTTGCA 3'

for first-strand cDNA synthesis. 5' RACE PCR was performed using an Abridged Anchor (AAP) primer (Gibco BRL, Grand Island, NY), corresponding to SEQ.

- 10 ID. No. 10, as follows:

5' GGCCACGCGTCGACTAGTACGGGIIIGGGIIIGG 3'

and GSP₂ primer, corresponding to SEQ. ID. No. 11, as follows:

5' TAACATGAACATGCAGGGAGTC 3'.

[0064] Nested PCR was performed using an Abridged Universal Anchor

- 15 primer (AUAP) (Gibco BRL, Grand Island, NY), corresponding to SEQ. ID. No. 12, as follows:

5' GGCCACGCGTCGACTAGTAC 3'

and GSP₃, corresponding to SEQ. ID. No. 13, as follows:

5' GGGAGTCGGAGATAGCATTG 3'.

- 20 After amplification, the PCR products were cloned into pCR[®] vector for subsequent characterization.

Example 4 - ORFX Has Homologs in Other Plant Species and Predicted Structural Similarity to Human Oncogene RAS Protein

[0065] Sequence analysis of *ORFX* revealed that it contains two introns and encodes a 163 amino acid polypeptide of approximately 22kD, shown in Figures 4A-B. Comparison of the predicted amino acid sequence of the *ORFX*

- 25 cDNA against sequences in the Genbank EST database found matches only with plant genes. Figures 4A-B show a CLUSTALW alignment of LpORFX (*L.pennellii*, AF261775, SEQ. ID. No. 2) and LeORFX (*L.esculentum*, AF261774,

LpORFX (SEQ. ID. No. 2) and LcORFX (SEQ. ID. No. 4) residues are shaded black when identical to at least 73% of all the genes included in the analysis.

[0066] As shown in Figures 4A-B, matches up to 70% similarity were found with ESTs in both monocotyledonous and dicotyledonous species. In addition, a weaker match (56.7% similarity) was found with a gymnosperm, *Pinus* (Pt)(SEQ. ID. No. 39). In tomato, at least four additional paralogs of *ORFX* were identified in the EST database. Eight homologs of *ORFX* appear in *Arabidopsis* genomic sequence, often in 2 or 3-gene clusters, and having intron-exon arrangements similar to *ORFX*. None of the putative homologs of *ORFX* has a known function. Thus, *ORFX* appears to represent a previously uncharacterized plant-specific multigene family.

[0067] Analysis of the predicted amino acid sequence indicates that ORFX is a soluble protein with α/β type secondary structure, shown in Figure 5A. Figure 5B shows the threading program LOOPP analysis, (predicted ORFX protein was compared to a training set of 594 structures, chosen from PDB to eliminate redundancy, using the LOOPP algorithms) assigns ORFX to the fold of 6q21, domain A, which is human oncogene RAS protein. The Z-scores for global

and local alignments of ORFX are 3.2 and 4, respectively, suggesting an overall shape similar to G-proteins. The detailed comparison of ORFX sequence with that of the RAX (where X can be S, N or D) family, reveals conserved fingerprints at RAX binding domains. The RAX family includes proteins with wide regulatory functions, including control of cell division (Sprang, S. R., "G Proteins, Effectors and GAPs: Structure and Mechanism," Curr. Opin. Struct. Biol. 7:849-56 (1997), which is hereby incorporated by reference in its entirety).

Example 5 - The Basis for Allelic Differences at *fw2.2*

[0068] In an effort to understand the basis for allelic differences at *fw2.2*, the *L. pennellii* and *L. esculentum* ORFX alleles were compared by amplifying and sequencing a 830 nt fragment containing ORFX (including 55 nt from the 3'UTR and 95 nt from the 5'UTR) from both NILs. Of the 42 nt differences between the two alleles, 35 fell within the two predicted introns, four represent silent mutations, and only three cause amino-acid changes. All three of the substitutions occurred within the first nine residues of the ORF, indicated as asterisks in Figure 4A. Although the start methionine cannot be determined with certainty, if the second methionine in the ORF, shown in Figure 5, were used, this would place all three potential substitutions in the 5' UTR. Conservation between the alleles suggests that the *fw2.2* phenotype is probably not caused by differences within the coding region of ORFX, but by one or more changes upstream in the promoter region of ORFX. Variation in upstream regulatory regions of the *teosinte branched1* gene has also been implicated in the domestication of maize (Wang et al., "The Limits of Selection During Maize Domestication," Nature 398:236-39 (1999), which is hereby incorporated by reference in its entirety). However, differences in fruit size imparted by the different *fw2.2* alleles may be modulated by a combination of sequence changes in the coding and upstream regions of ORFX (Phillips, P.C., "From Complex Traits to Complex Alleles," Trends in Genetics 15: 6-8 (1999), which is hereby incorporated by reference in its entirety).

[0069] A reduction in cell division in carpels of the small-fruited NIL is correlated with overall higher levels of ORFX transcript, suggesting that ORFX may be a negative regulator of cell division. Whether the ORFX and RAX proteins share common properties other than predicted 3D structure and control of

cell division awaits future experimentation. An affirmative result may reflect an ancient and common origin in processes of cell cycle regulation in plants and animals.

- [0070]** Although the invention has been described in detail for the purpose of illustration, it is understood that such detail is solely for that purpose, and variations can be made therein by those skilled in the art without departing from the spirit and scope of the invention which is defined by the following claims.
- 5

0598659-070001